United Nations ECE/CES/2015/43

Economic and Social Council

Distr.: General
1 June 2015

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-third plenary session
Geneva, 15-17 June 2015
Item 4 of the provisional agenda
Modernization of statistical production and services and managing for efficiency

Annual report of the High-Level Group for the Modernization of Statistical Production and Services

Note by the High-Level Group

Summary

This paper informs the Conference of European Statisticians about the activities and outputs of the High-Level Group for the Modernization of Statistical Production and Services. The purpose of the Group is to guide strategic developments in the ways official statistics are produced.

The Conference of European Statisticians is invited to express its views on the value of the work done so far and provide guidance on possible future priorities.

GE.15-11047 (E)

1511047

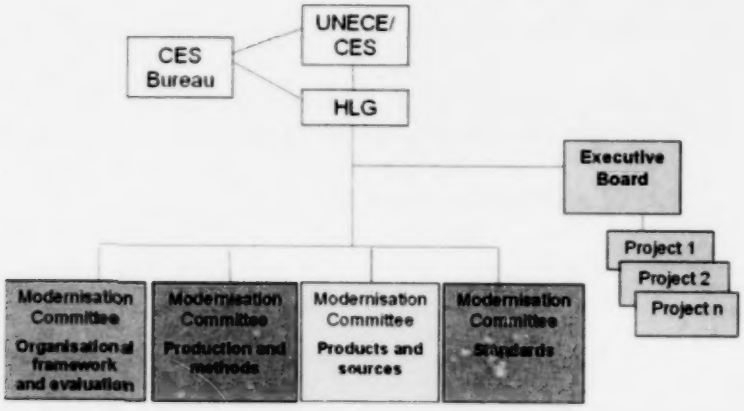
Please recycle

I. Introduction

1. The High-level Group for the Modernization of Statistical Production and Services (HLG) was created by the Bureau of the Conference of European Statisticians in 2010. It comprises the heads of eleven national and international statistical organizations, and has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced.

2. To further the work on statistical modernization, the HLG oversees four modernization committees, as well as annual multi-national collaboration projects, which are managed by an Executive Board. Figure 1 illustrates this governance structure, which was endorsed by the Conference of European Statisticians in 2013.

Figure 1
Governance structure for HLG activities



3. To determine the topics for the projects, the HLG organises a workshop each November, inviting representatives of various expert groups and projects related to modernization of official statistics. These workshops review progress and determine the key priorities for the following year. In November 2013, the workshop decided that implementing the Common Statistical Production Architecture (CSPA)[1] and investigating the potential use of Big Data sources for official statistics were the two highest priorities for 2014. In November 2014, the priorities decided for 2015 were further work on the implementation of the Common Statistical Production Architecture, and development of the Big Data "Sandbox" for the production of "real" statistics.

II. Common Statistical Production Architecture

4. The Common Statistical Production Architecture (CSPA) was created during 2013 in an international collaboration project under HLG. That project included the

[1] For more information about how the CSPA supports the modernization of official statistics by facilitating the concept of "plug and play" for the components of statistical production, please see http://www1.unece.org/stat/platform/display/CSPA

development and specification of the architecture, as well as a practical test of its principles and applicability in a "proof of concept". However, given the strict time limits for that project (one year), it deliberately did not include the implementation of the architecture for production systems within statistical organizations.

5. CSPA provides a standard framework for developing the components of statistical production in a way that they can be much more easily shared within and between organizations. It builds on key international standards such as the Generic Statistical Business Process Model (GSBPM), and the Generic Statistical Information Model (GSIM).

6. During the 2014 project the following CSPA-compliant services were developed:

- Seasonal adjustment – France, Australia, New Zealand
- Confidentialized analysis of microdata – Canada, Australia
- Statistical chart generator – OECD
- SDMX transformer – OECD
- Sample selection – Netherlands
- Linear error localisation – Netherlands
- Linear rule checking – Netherlands
- Error correction – Italy

7. These services have been incorporated into a "service catalogue", which is hosted by Eurostat, and are available for use by other statistical organizations. Another result of the 2014 project was a slightly revised version of CSPA (version 1.1), which incorporates additional guidance based on the lessons learned from the project.

8. The 2015 project has the following objectives:

- To extend the governance and support offered for the implementation of CSPA compliant statistical services;
- To build more CSPA services and have a plan in place for continued investment in the development of CSPA services;
- To facilitate the transitioning of CSPA governance to the Modernization Committee on Production and Methods.

9. To achieve these objectives, work is currently focusing on:

- Developing a common investment strategy for participating organizations based on greater alignment of business plans and identifying common capability requirements;
- Creating a CSPA Logical Information Model to facilitate service specification and implementation;
- Developing new CSPA services, e.g. for probabilistic record linkage, classification management, and validation rules specification.

III. Big Data

10. The HLG has identified the need to better understand the impact of new data sources, and in particular "Big Data" on official statistics. Following the

endorsement of the HLG paper "What does Big Data mean for official statistics?" (ECE/CES/2013/20) by the Conference of European Statisticians, a project was launched in 2014, with three main objectives:

(a) To identify, examine and provide guidance for statistical organizations on the main possibilities offered by Big Data and to act upon the main strategic and methodological issues that Big Data poses for the official statistics industry;

(b) To demonstrate the feasibility of efficient production of both novel products and 'mainstream' official statistics using Big Data sources, and the possibility to replicate these approaches across different national contexts;

(c) To facilitate the sharing across organizations of knowledge, expertise, tools and methods for the production of statistics using Big Data sources.

11. The main outcomes of the 2014 project were:

(a) A framework for assessing the quality of Big Data;

(b) Papers on the current status of statistical disclosure control, the characteristics of Big Data and their implications for data privacy, and practical measures to manage Big Data privacy;

(c) Guidelines for the establishment of partnerships in Big Data projects for official statistics;

(d) A survey on skills and training needed to work with Big Data, and a draft skills profile for data scientists;

(e) The "Sandbox", a state of the art web-accessible computer environment, created in partnership with the Irish Central Statistics Office and the Irish Centre for High-End Computing, which provides a platform for collaboration where participants can experiment with different types of Big Data, and learn how to use the latest software tools for Big Data processing.

12. All results of the 2014 project, including the outcomes of the initial Sandbox experiments are available at:

http://www1.unece.org/stat/platform/display/bigdata/2014+Project

13. The 2015 project is focusing much more on the use of the Sandbox as an international resource for Big Data experiments and training, though it also has the very concrete goal of producing a set of internationally comparable statistics from one or more Big Data sources by the end of 2015. The scope of the project was deliberately defined to ensure coordination with related projects being led by Eurostat and the UN Statistical Division, who are both active partners in the HLG Big Data work. Current work includes investigation of data from Twitter and Wikipedia, as well as potential efficiency savings using Big Data tools to process traditional sources such as the UN Statistical Division foreign trade data.

14. The value of the Sandbox as a resource for the international statistical community is being increasingly recognised. A "sprint" session will be held in Cork, Ireland, on 22-24 June, bringing together representatives of key national and international statistical organizations, to develop proposals on the future governance and funding model for the Sandbox.

IV. The work of the modernization committees

A. Modernization Committee on Organization Framework and Evaluation

15. Key items of work completed or currently in progress within the modernization committee include:

- Survey on best practices in change and risk management;
- Study on legal and licensing issues, which led to proposals for a statement of intent;
- Generic skills profile for staff working with Big Data in statistical organizations is currently being tested;
- Guidelines for managers, including best practices, under development.

B. Modernization Committee on Production and Methods

16. Key items of work completed or currently in progress within the modernization committee include:

- Paper on the state of the art in the use of machine learning in statistical organizations;
- Generic Framework for Statistical Data Editing – draft prepared for peer review;
- Modernization maturity model - initial work to develop a model to assess the state of readiness of a statistical organization to adopt HLG outputs.

C. Modernization Committee on Products and Sources

17. Key items of work completed or currently in progress within the modernization committee include:

- Inventory of Big Data activities in statistical organizations;
- Study on the use of mobile devices for data collection and dissemination;
- Papers and survey of best practices on marketing and communicating the value of official statistics;
- Best practices in linked open data, including formalising and publishing key metadata in linked open data formats.

D. Modernization Committee on Standards

18. Key items of work completed or currently in progress within the modernization committee include:

- Generic Activity Model for Statistical Organizations (GAMSO) – Version 1.0 released in March 2015;
- Glossary of metadata / modernization terminology - under development;

- Quality indicators for each sub-process of the Generic Statistical Business Process Model – first version released for public consultation.

19. In addition to the four modernization committees, the HLG has established a cross-cutting task team on communicating modernization, which is looking at how to improve the flow of knowledge about modernization activities within the official statistics community. It will report at the end of 2015.

V. Ensuring coordination

20. All of the activities listed above are carried out in close cooperation with other international organizations and initiatives. For example, Eurostat and OECD are members of the HLG. Eurostat staff are involved in almost all of the activities mentioned, whilst OECD, the European Central Bank, the UN Statistical Division, the World Bank, the World Trade Organization, and the UN Industrial Development Organization (UNIDO) are involved in those activities that are relevant to their work. There is close cooperation with the UN Economic and Social Commission for Asia and the Pacific, which has established mechanisms to spread information about modernization activities to its member countries.

21. These activities are not limited to those working in official statistics. Where relevant, partnerships are also forged with external bodies. For example the Data Documentation Initiative and academia are represented on the modernization committees, whilst the partnerships with the Irish Centre for High-End Computing, data providers and software companies have been essential to the success of the Big Data projects.

VI. Looking forwards

22. There are several recent developments that could influence the priorities for the HLG in 2016 and beyond:

(a) The "Statistical Modernization Community" – a proposal to put institutional participation in modernization activities on a more formal basis by introducing a "statement of intent" which organizations can sign up to. In many respects, this would simply formalise the current situation, with a clearer demonstration of support for the activities of the HLG;

(b) Developing the notion of "competencies" as the key "units" for investment planning and collaborative development. Competencies are defined as things that an organization does, or wants to do. They are combinations of "factors of production" such as methods, technology, people, skills and standards. An example of a statistical competency is seasonal adjustment. Joint work with the European Statistical System Task Force on Enterprise Architecture is foreseen to further elaborate a competency framework;

(c) Increased sharing and coordination of investment planning. If statistical organizations have easy access to information on who is investing, or planning to invest in improving capabilities in specific areas, formal and informal collaboration opportunities should be much easier to identify;

(d) Data integration. The work on Big Data and other new data sources (and modes of collection) shows that data integration from multiple sources and mixed collection modes will become increasingly important for statistical organizations over the next few years. Developing and enhancing data integration

techniques, methods and tools could be the subject of an international collaboration project in 2016;

(e) Reporting. The UN Statistical Commission decided in March 2015 to request regular reports on statistical modernization work from the Conference of European Statisticians. This will increase the global outreach of HLG activities.

VII. Input required from the Conference of European Statisticians

23. The Conference is invited to express its views on the value of the work done so far under the HLG and provide guidance on possible future priorities.